

Optimal DoF of the K-User Broadcast Channel with Delayed and Imperfect Current CSIT

Paul de Kerret, David Gesbert, Jingjing Zhang, and Petros Elia

Communication Systems Department,

EURECOM

Abstract

This work¹ studies the optimal Degrees-of-Freedom (DoF) of the K -User MISO Broadcast Channel (BC) with delayed Channel-State Information at the Transmitter (CSIT) and with additional current noisy CSIT where the current channel estimation error scales in $P^{-\alpha}$ for $\alpha \in [0, 1]$. This paper establishes for the first time the optimal DoF in this setting thanks to a new transmission scheme which achieves the elusive DoF-optimal combining of the Maddah-Ali and Tse scheme (MAT) introduced in their seminal work in 2010 with Zero-Forcing (ZF) for an arbitrary number of users. The derived sum DoF takes the surprisingly simple form $(1 - \alpha)K/H_K + \alpha K$ where $H_K \triangleq \sum_{k=1}^K \frac{1}{k}$ is the sum-DoF achieved using solely MAT.

I. INTRODUCTION

In the K -user wireless BC, feedback accuracy and timeliness crucially affects performance, but are also notoriously difficult to obtain. In terms of accuracy, it is well known that increasing feedback quality can elevate performance, from that of TDMA (sum DoF of 1), to the maximum possible interference-free performance with a sum DoF of K . As the recent result in [1] tells us, having imperfect instantaneous CSIT with an estimation error that scales (in the high-power P setting) as $P^{-\alpha}$ ($\alpha \in [0, 1]$), can allow, using basic ZF precoding techniques and rate splitting, for an optimal sum-DoF of $1 + (K - 1)\alpha$.

¹D. Gesbert and P. de Kerret are supported by the European Research Council under the European Union's Horizon 2020 research and innovation program (Agreement no. 670896). P. Elia is supported by the ANR project ECOLOGICAL-BITS-AND-FLOPS.

Parts of these results have been published in IEEE Information Theory Workshop 2016, Cambridge.

On the other hand, when perfect-accuracy CSIT is obtained in a delayed manner — in the sense that the CSI is fed back with a delay exceeding the channel coherence period — then, using more involved, retrospective, *MAT*-type space-time alignment [2], one can surprisingly get substantial DoF gains, reaching a sum-DoF of K/H_K with $H_K \triangleq \sum_{k=1}^K \frac{1}{k}$, which scales with K approximately as $K/\ln(K)$.

This interplay between performance and feedback timeliness-and-quality, has sparked a plethora of works that considered a variety of feedback mechanisms with delayed and imperfect CSIT. Such works can be found in [3]–[5], and in [6] which — for the two-user MISO BC setting — studied the case where the CSIT can alternate between perfect, delayed (completely outdated), and non-existent (see also [7]). The performance at finite SNR of the MAT scheme were also discussed in [8], while the gap to optimality was bounded in [9].

An interesting approach came with the work in [10] which introduced a feedback scenario that offered a combination of imperfect-quality current (instantaneously available) CSIT, together with additional (perfect-accuracy) delayed CSIT. In this same setting — which reflected different applications, including that of using predictions to get an estimate of the current state of a time-correlated channel — the channel estimation error of the current channel state was assumed to scale in power as $P^{-\alpha}$, for some CSIT quality exponent $\alpha \geq 0$. [10] also introduced ingredients that proved to be key in this setting: The use of an initial imperfect ZF precoding layer, followed by retransmission of a *quantized* form of the interference generated by CSI imperfections in the first layer precoder.

Additional work — within the context of the BC — came in [11], [12] which established the maximal DoF in a two-user MISO BC scenario, as well as in [13], [14] which considered the case of imperfect-quality delayed CSIT. More results can also be found in [14], [15] which considered the broad setting of any-time any-quality feedback, and in [16], [17] which studied the two-user MIMO BC (and IC); all for the two-user case. This general challenge of dealing with imperfect feedback has also sparked very recent interest, with different publications that include [18]–[23]. Finally, interesting connections between the delayed CSIT configuration and the generalized feedback setting [24] has been put forward in [25].

A. Simultaneous scaling of MAT and ZF gains

For the more general case of the K -user BC, again with joint delayed and imperfect-current CSIT, very little is known. For the particular case considered here, a general outer bound was provided in [26], and efforts to reach this bound can be found in [27]. The main goal has remained to secure simultaneous scaling of MAT-type gains (that exploit delayed CSIT), and ZF gains (that exploit imperfect-quality current CSI). To date, this has remained an elusive open problem, and any instance of providing such simultaneous gains was either limited to the 2-user case, or — as in the case of the scheme in [26] — resulted in MAT-type DoF gains that saturated at 2. This elusive open problem is resolved here, by inventing a new scheme, referred to as the Q-MAT scheme, that combines different new ingredients that jointly allow for MAT and ZF components to optimally coexist. Combined with the outer bound in [26], the achieved DoF establishes the optimal sum-DoF, which is here shown to be equal to $\alpha K + (1 - \alpha)K / (\sum_{k=1}^K \frac{1}{k})$.

B. Notation

For $\mathcal{C}(P)$ denoting the sum capacity [28] of the MISO BC considered, we will place emphasis on the high-SNR degree of freedom approximation

$$\text{DoF}^* \triangleq \lim_{P \rightarrow \infty} \frac{\mathcal{C}(P)}{\log_2(P)}. \quad (1)$$

We will use the notation $H_K \triangleq \sum_{i=1}^K \frac{1}{i}$ to represent the K -th harmonic number. \mathbb{Z} will represent the integers, \mathbb{Z}^+ the positive integers, \mathbb{R}^+ the positive real numbers, \mathbb{C} the complex numbers, $\binom{n}{k}$ the n -choose- k operator, and \oplus the bitwise XOR operation. We will use $\mathcal{K} \triangleq \{1, 2, \dots, K\}$. If $\mathcal{S} \subset \mathcal{K}$ is a set, then $\bar{\mathcal{S}}$ will denote $\mathcal{K} \setminus \mathcal{S}$, and $|\mathcal{S}|$ will denote its cardinality. Complex vectors will be denoted by lower-case bold font. For any vector \mathbf{x} , we will use $\{\mathbf{x}\}_i$, $\|\mathbf{x}\|^2$ and \mathbf{x}^H to respectively denote the i th element of the vector, its magnitude-squared, and its conjugate transpose. We will also use \doteq to denote *exponential equality*, i.e., we write $f(P) \doteq P^B$ to denote $\lim_{P \rightarrow \infty} \frac{\log(f(P))}{\log(P)} = B$. We write $\mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ to denote the complex Gaussian distribution of zero mean and variance σ^2 .

C. System Model

1) *K-User MISO Broadcast Channel*: This work considers the K -User MISO BC with fading, where the Transmitter (TX) — which is equipped with M antennas ($M \geq K$) — serves K single-

antenna Receivers (RXs). At any time t , the signal received at RX $k \in \mathcal{K}$, can be written as

$$y_k[t] = \mathbf{h}_k^H[t] \mathbf{x}[t] + n_k[t] \quad (2)$$

where $\mathbf{h}_k^H[t] \in \mathbb{C}^{1 \times M}$ represents the channel to user k at time t , where $\mathbf{x}[t] \in \mathbb{C}^M$ represents the transmitted signal, and where $n_k[t] \in \mathbb{C}$ represents the additive noise at RX k , where this noise is distributed as $\mathcal{N}_{\mathbb{C}}(0, 1)$, independently of the channel and of the transmitted signal. Furthermore, the transmitted signal $\mathbf{x}[t]$ fulfills the average asymptotic power constraint $\mathbb{E}[\|\mathbf{x}[t]\|^2] \doteq P$. The channel is assumed to be drawn from a continuous ergodic distribution such that all the channel matrices and all their sub-matrices are almost surely full rank.

2) *Perfect delayed CSIT and imperfect current CSIT*: Our CSIT model builds on the delayed CSIT model introduced in [2] and generalized in [10]–[12] to account for the availability of an imperfect estimate of the current channel state. For ease of exposition, we will here adopt the fast-fading channel model, and will assume that at any time t , the TX has access to the delayed CSI (with perfect accuracy) of all previous channel realizations up to time $t - 1$, as well as an imperfect estimate of the current channel state. Each current estimate $\hat{\mathbf{h}}_k^H[t]$ for each channel $\mathbf{h}_k^H[t]$, comes with an estimation error

$$\tilde{\mathbf{h}}_k^H[t] = \mathbf{h}_k^H[t] - \hat{\mathbf{h}}_k^H[t] \quad (3)$$

whose entries are i.i.d. $\mathcal{N}_{\mathbb{C}}(0, P^{-\alpha})$ with power $P^{-\alpha}$ for some parameter $\alpha \in [0, 1]$, which we refer to as the *CSIT quality exponent*, and which is used to parameterize the accuracy of the current CSIT². All estimates are assumed to be independent of all estimation errors. Finally we make the common assumption that the channel $\mathbf{h}_k^H[t]$ is independent of all previous channel estimates and channel estimation errors, when conditioned on $\hat{\mathbf{h}}_k^H[t]$. We also adhere to the common convention (see [2], [10], [12]) of assuming perfect and global knowledge of channel state information at the receivers (perfect global CSIR), where the receivers know all channel states and all estimates.

²Note that from a DoF perspective, we can restrict ourselves to $\alpha \in [0, 1]$, since an estimation/quantization error with power scaling as P^{-1} ($\alpha = 1$), is essentially perfect. Similarly an estimation error with power scaling as P^0 ($\alpha = 0$), offers no DoF gains over the case of having no CSIT (cf. [1], see also [26]).

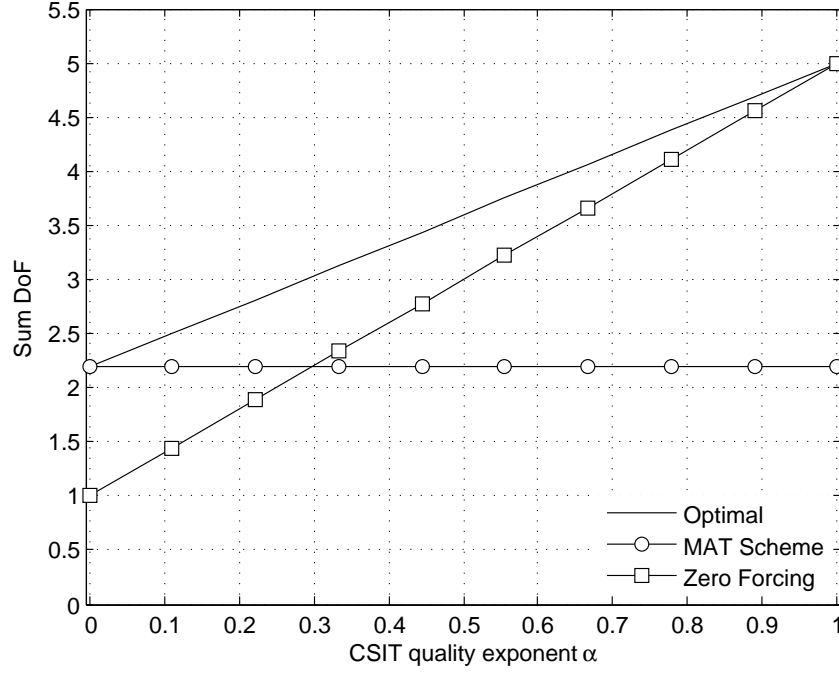


Fig. 1: Sum DoF achieved as a function of the CSIT scaling exponent α for $K = 5$ users.

II. MAIN RESULTS

We proceed directly with the main result.

Theorem 1. *In the K -user MISO BC ($M \geq K$) with perfect delayed CSIT and α -quality current CSIT, the optimal sum DoF is*

$$\text{DoF}^*(\alpha) = (1 - \alpha) \frac{K}{\sum_{k=1}^K \frac{1}{k}} + \alpha K. \quad (4)$$

Proof: The scheme that achieves this DoF, is described in the next sections. The optimality follows from the fact that this scheme's performance matches the DoF outer bound in [26]. ■

Figure 1 shows the optimal DoF for the 5-user case, and compares this performance to the MAT-only and ZF-only DoF performance. The proposed scheme significantly outperforms both ZF and MAT, and the gap increases with K .

We proceed by describing first the simpler case of $K = 2$ so as to get the proper insight to understand the general K case, presented after.

III. SCHEME FOR THE $K = 2$ USER CASE

As stated above, we begin with the scheme description for the two-user case. Although the main motivation for the our scheme is clearly the case of $K > 2$, this simple setting contains all the key components and allows to present in detail all the steps, with lighter notations than in the general K -user case. Note that the $K = 3$ -user case is described in the conference version of this work [29] and can serve as an additional example to gain insights before dealing with the general case ³.

A. Encoding and Transmission

The transmission follows closely the MAT multi-destination multi-layer scheme in [2] and we refer to this work for a nice and intuitive description of the MAT scheme. Consequently, our scheme is also divided into two phases for the 2-user case, where phase 1 corresponds to the transmission of order-1 data symbols (meant for one user at a time) and spans 2 Time Slots (TS), while phase 2 corresponds to the transmission of order-2 data symbols (meant for both users simultaneously) and spans 1 TS.

A first deviation from existing schemes can be found in the fact that the scheme requires several *rounds* (each of them following the structure in phases of the MAT scheme).

Remark 1. The encoding across rounds was already used in [26] and is a consequence of the delayed CSIT assumption. Indeed, the CSI necessary for the transmission of some of the data symbols (so-called *auxiliary data symbols*) is not available when the transmission occurs. Therefore, this auxiliary data symbol is instead transmitted in the same phase of the next round. \square

We describe the transmission for an arbitrary round N , and we will use the shorthand notation $\bullet^{(R_N)}$ to denote the fact that the index of the round is equal to N . The particularities of the first and last rounds will be clarified later on.

³While a bit premature at this stage, we hasten to note that this exposition of the two-user case, nicely accentuates the difference between our scheme and previous efforts [11], [12] which — albeit optimal for the two user case — had to deviate from the MAT-type canonical structure in order to accommodate for the ZF component, thus making it difficult to extend to higher dimensions.

1) *Phase 1*: Phase 1 spans two time slots, denoted by $t_1^{(R_N)}$ and $t_2^{(R_N)}$. For $t = t_1^{(R_N)}$, the transmitted signal is given by

$$\mathbf{x}[t] = \mathbf{V}[t]\mathbf{m}[t] + \mathbf{v}_2^{\text{ZF}}[t]a_2[t] + \sum_{k=1}^2 \mathbf{v}_k^{\text{ZF}}[t]s_k[t] \quad (5)$$

where

- $\mathbf{m}[t] \in \mathbb{C}^2$ is a vector containing two so-called *Q-MAT data symbols* meant for user 1, each carrying $(1-\alpha)\log_2(P)$ bits, where the first symbol is allocated full power $\mathbb{E}[|\{\mathbf{m}[t]\}_1|^2] \doteq P$, while the second symbol is allocated lesser power $\mathbb{E}[|\{\mathbf{m}[t]\}_2|^2] \doteq P^{1-\alpha}$. Furthermore $\mathbf{V}[t] \in \mathbb{C}^{M \times 2}$ is defined as

$$\mathbf{V}[t] \triangleq \begin{bmatrix} \mathbf{v}_1^{\text{ZF}}[t] & \mathbf{u}_1 \end{bmatrix} \quad (6)$$

where $\mathbf{v}_1^{\text{ZF}}[t] \in \mathbb{C}^M$ is the unit-norm ZF beamformer aimed at user 1 (i.e., which is orthogonal to the current estimate of the channel to user 2), while $\mathbf{u}_1 \in \mathbb{C}^M$ is a unit-norm vector that is randomly drawn and isotropically distributed.

- $a_2[t] \in \mathbb{C}$ is a so-called *auxiliary data symbol* meant for user 2, carrying $\min(1-\alpha, \alpha)\log_2(P)$ bits (generally from previous interfering terms), and allocated full power $\mathbb{E}[|a_2[t]|^2] \doteq P$.
- $s_k[t] \in \mathbb{C}$, $k \in \{1, 2\}$ are *ZF data symbols* meant for user k , each carrying $\alpha\log_2(P)$ bits and each having power $\mathbb{E}[|s_k[t]|^2] \doteq P^\alpha$.

Upon omitting the noise realizations for simplicity, the received symbols during $t = t_1^{(R_N)}$, can be written as

$$\begin{aligned} y_1[t] &= \underbrace{\mathbf{h}_1^H[t]\mathbf{V}[t]\mathbf{m}[t]}_{\doteq P} + \underbrace{\mathbf{h}_1^H[t]\mathbf{v}_2^{\text{ZF}}[t]a_2[t]}_{\doteq P^{1-\alpha}} + \underbrace{z_1[t]}_{\doteq P^\alpha} \\ y_2[t] &= \underbrace{\mathbf{h}_2^H[t]\mathbf{v}_2^{\text{ZF}}[t]a_2[t]}_{\doteq P} + \underbrace{i_2[t]}_{\doteq P^{1-\alpha}} + \underbrace{z_2[t]}_{\doteq P^\alpha} \end{aligned} \quad (7)$$

where

$$i_2[t] \triangleq \underbrace{\mathbf{h}_2^H[t]\mathbf{V}[t]\mathbf{m}[t]}_{\doteq P^{1-\alpha}} \quad (8)$$

and where for any $t \in \mathbb{Z}^+$,

$$z_k[t] \triangleq \underbrace{\mathbf{h}_k^H[t]\mathbf{v}_k^{\text{ZF}}[t]s_k[t]}_{\doteq P^\alpha} + \underbrace{\mathbf{h}_k^H[t]\mathbf{v}_k^{\text{ZF}}[t]s_{\bar{k}}[t]}_{\doteq P^0}, \quad k \in \{1, 2\}. \quad (9)$$

In the above equations, underneath each summand, we describe the asymptotic approximation of the power of the corresponding term.

For $t = t_2^{(R_N)}$, the transmission is described by the above equations, after exchanging the indices of the two users ($1 \leftrightarrow 2$). Thus, the received signals during $t = t_2^{(R_N)}$, take the form

$$\begin{aligned} y_1[t] &= \underbrace{\mathbf{h}_1^H[t] \mathbf{v}_1^{\text{ZF}}[t] a_1[t]}_{\doteq P} + \underbrace{i_1[t]}_{\doteq P^{1-\alpha}} + \underbrace{z_1[t]}_{\doteq P^\alpha} \\ y_2[t] &= \underbrace{\mathbf{h}_2^H[t] \mathbf{V}[t] \mathbf{m}[t]}_{\doteq P} + \underbrace{\mathbf{h}_2^H[t] \mathbf{v}_1^{\text{ZF}}[t] a_1[t]}_{\doteq P^{1-\alpha}} + \underbrace{z_2[t]}_{\doteq P^\alpha} \end{aligned} \quad (10)$$

where now

$$i_1[t] \triangleq \underbrace{\mathbf{h}_1^H[t] \mathbf{V}[t] \mathbf{m}[t]}_{\doteq P^{1-\alpha}}. \quad (11)$$

a) Interference quantization: At the end of phase 1, the TX can use its delayed CSIT to compute the interference terms $i_2[t_1^{(R_N)}]$ and $i_1[t_2^{(R_N)}]$. These terms are first quantized according to a specifically designed quantizer, before being used to generate the data symbols which will be transmitted either in the next phase of the same round or in the same phase of the next round.

We now present the quantization scheme, which needs to be differentiated depending whether $\alpha \leq \frac{1}{2}$ or not. Let us first consider the more involved case of $\alpha \leq \frac{1}{2}$. Our quantizer is defined in the following lemma.

Lemma 1. *Let Y be a unit-variance zero-mean random variable of bounded density p_Y . Then, there exists a quantizer of rate $(\beta_1 - \beta_2) \log_2(P)$ bits for any $0 < \beta_2 \leq \beta_1$, denoted by Q_{β_1, β_2} , such that, for any unit-variance zero-mean random variable n , it holds that*

$$\lim_{P \rightarrow \infty} \Pr\{Q_{\beta_1, \beta_2}(\sqrt{P^{\beta_1}}y + \sqrt{P^{\beta_2}}n) = Q_{\beta_1, \beta_2}(\sqrt{P^{\beta_1}}y)\} = 1 \quad (12)$$

and at the same time

$$\mathbb{E} \left[|Q_{\beta_1, \beta_2}(\sqrt{P^{\beta_1}}y) - \sqrt{P^{\beta_1}}y|^2 \right] \leq P^{\beta_2}. \quad (13)$$

Proof: The design of the quantizer and the detailed proof of the result is given in the appendix. ■

To convey some intuition on the particular properties of this quantizer, we now describe how it could be used in the following toy-example.

Example 1. *Let us consider a setting where a first node, node A , has the knowledge of a Gaussian random variable $X \sim \mathcal{N}_{\mathbb{C}}(0, P)$ while another node, node b only obtains a corrupted*

version Y given by

$$Y = \underbrace{X}_{\doteq P} + \underbrace{N}_{\doteq P^\gamma} \quad (14)$$

where N is a Gaussian random variable with variance P^γ . Node A wants to transmit the minimum number of bits to node B in order for this node to reconstruct the random variable X up to the noise floor, as the power P increases. As a solution to this problem, we propose to use the quantizer $Q_{1,\gamma}$ described in the previous lemma. Indeed, using this quantizer, it holds that

$$\lim_{P \rightarrow \infty} \Pr\{Q_{1,\gamma}(X + N) = Q_{1,\gamma}(X)\} = 1 \quad (15)$$

where

$$X = Q_{1,\gamma}(X) + n_Q \quad (16)$$

and with $n_Q \doteq P^\gamma$. Hence, if node A quantizes n_Q using $\gamma \log_2(P)$ bits using an adequate quantizer from the literature, and transmit it to node B , it follows from well-known results from Rate-Distortion theory that node B will be able to reconstruct X up to the noise floor, as the power P increases. Indeed, $Q_{1,\gamma}(X)$ can be obtained with probability one from $X + N$, as P increases. \square

Let us now consider the quantization of the interference term $i_2[t_1^{(R_N)}]$ (scaling in $P^{1-\alpha}$). Using the above quantizer Q_{β_1, β_2} , and setting $\beta_1 = 1 - \alpha, \beta_2 = \alpha$, offers us a quantized version $Q_{1-\alpha, \alpha}(i_2[t_1^{(R_N)}])$ that carries $(1 - 2\alpha) \log_2(P)$ bits, and which guarantees that the resulting quantization noise, which we denote by $n_2[t_1^{(R_N)}]$, has a power that scales as P^α . The aforementioned quantization noise $n_2[t_1^{(R_N)}]$ is then itself re-quantized using any standard optimal quantizer with $\alpha \log_2(P)$ bits, which is known [28] to guarantee quantization noise (from the second quantization) that scales as P^0 . For $\hat{n}_2[t_1^{(R_N)}]$ denoting the quantized version of $n_2[t_1^{(R_N)}]$, we get the final combined estimate carrying $(1 - \alpha) \log_2(P)$ bits in the form

$$\hat{i}_2[t_1^{(R_N)}] \triangleq Q_{1-\alpha, \alpha}(i_2[t_1^{(R_N)}]) + \hat{n}_2[t_1^{(R_N)}] \quad (17)$$

where in the above the addition is over the complex numbers. This 2-step quantization is illustrated in Fig. 2.

For the easier case where $\alpha \geq \frac{1}{2}$, the interference term $i_1[t_2^{(R_N)}]$ is simply quantized using $(1 - \alpha) \log_2(P)$ bits using a standard quantizer, guaranteed (cf. [28]) to have quantization noise that scales in P^0 . The quantized signal obtained is also denoted by $\hat{i}_2[t_1^{(R_N)}]$.

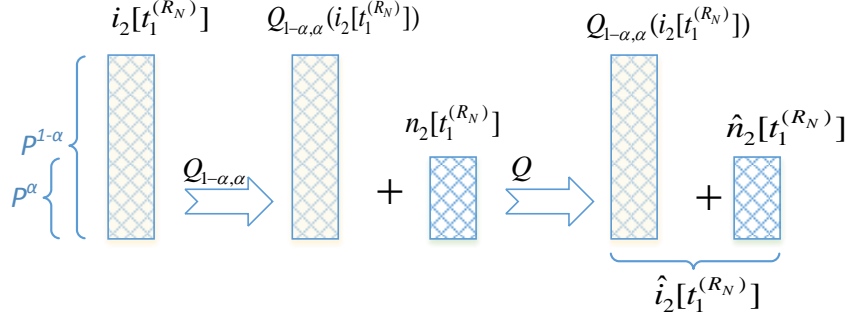


Fig. 2: Illustration of the 2-step quantization scheme with the quantization noise at the noise floor being omitted.

The same quantization process is applied to $i_1[t_2^{(R_N)}]$ to obtain $\hat{i}_1[t_2^{(R_N)}]$ with the same rate and quantization-noise properties. These quantized bits will be placed in the auxiliary data symbols of the next round, as we now describe.

b) Generation of auxiliary data symbols for phase 1 of round $N + 1$: The auxiliary data symbol $a_2[t_1^{(R_{N+1})}]$ will carry

$$\begin{cases} a_2[t_1^{(R_{N+1})}] \leftarrow \hat{n}_2[t_1^{(R_N)}] & , \alpha \leq \frac{1}{2} \\ a_2[t_1^{(R_{N+1})}] \leftarrow \hat{i}_2[t_1^{(R_N)}] & , \alpha \geq \frac{1}{2} \end{cases} \quad (18)$$

and similarly $a_1[t_2^{(R_{N+1})}]$ will carry the following quantization bits

$$\begin{cases} a_1[t_2^{(R_{N+1})}] \leftarrow \hat{n}_1[t_2^{(R_N)}] & , \alpha \leq \frac{1}{2} \\ a_1[t_2^{(R_{N+1})}] \leftarrow \hat{i}_1[t_2^{(R_N)}] & , \alpha \geq \frac{1}{2}. \end{cases} \quad (19)$$

During the first round, all these auxiliary symbols are initialized to zero.

c) Generation of Q-MAT data symbols for phase 2 of round N : During the second phase of round N , using delayed CSIT, the TX generates the following order-2 Q-MAT data symbol

$$m[t_{1,2}^{(R_N)}] \leftarrow (\hat{i}_2[t_1^{(R_N)}] \oplus \hat{i}_1[t_2^{(R_N)}]) \quad (20)$$

where $t_{1,2}^{(R_N)}$ denotes the TS used to transmit the order-2 data symbols.

2) *Phase 2:* Phase 2 consists of one TS $t = t_{1,2}^{(R_N)}$, during which

$$\mathbf{x}[t] = \mathbf{v}[t]m[t] + \sum_{k=1}^2 \mathbf{v}_k^{\text{ZF}}[t]s_k[t] \quad (21)$$

where

- $m[t] \in \mathbb{C}$ is an order-2 Q-MAT data symbol which carries $(1-\alpha) \log_2(P)$ bits that originate from the previous phase of the same round (see (20)), and which is destined for both users. The symbol is allocated full power $\mathbb{E}[|m[t]|^2] \doteq P$.
- $s_k[t]$, $k \in \{1, 2\}$ is a ZF data symbol destined for user k , carrying $\alpha \log_2(P)$ bits, and having power $\mathbb{E}[|s_k[t]|^2] \doteq P^\alpha$.

Upon omitting the noise realizations, the received signals during TS $t = t_{1,2}^{(R_N)}$ take the form

$$\begin{aligned} y_1[t] &= \underbrace{\mathbf{h}_1^H[t] \mathbf{v}[t] m[t]}_{\doteq P} + \underbrace{z_1[t]}_{\doteq P^\alpha} \\ y_2[t] &= \underbrace{\mathbf{h}_2^H[t] \mathbf{v}[t] m[t]}_{\doteq P} + \underbrace{z_2[t]}_{\doteq P^\alpha}. \end{aligned} \tag{22}$$

B. Decoding (achievability proof by induction)

We now turn to the decoding part, which here — for the sake of clarity — will be assumed to start after the end of transmission in all rounds and all phases. We will show that each user can decode all its desired data symbols. The proof has to be done by induction due to the fact that the auxiliary data symbols contain information coming from the previous round.

Let us consider without loss of generality the decoding at user 1. Our induction statement is that if the auxiliary data symbol $a_1[t_2^{(R_N)}]$ and $a_2[t_1^{(R_N)}]$ are decoded at user 1, user 1 can decode:

- Its destined Q-MAT data symbols $\mathbf{m}[t_1^{(R_N)}]$ and $m[t_{1,2}^{(R_N)}]$
- Its destined ZF data symbols $s_1[t_1^{(R_N)}]$, $s_1[t_2^{(R_N)}]$, $s_1[t_{1,2}^{(R_N)}]$
- The auxiliary data symbols of the following round $a_1[t_2^{(R_{N+1})}]$ and $a_2[t_1^{(R_{N+1})}]$, thus allowing for the inductive step.

The initialization is done for $n = 0$ by considering that all data symbols of round $n = 0$ have zero rate and by setting all auxiliary data symbols of round $n = 1$ to zero. Indeed, all data symbols for $n = 0$ are decoded, as well as the auxiliary data symbols of round $n + 1 = 1$. Thus, the induction property is satisfied for $n = 0$.

We then proceed to consider an arbitrary round $N > 0$. As part of the induction, we consider $a_1[t_2^{(R_N)}]$ and $a_2[t_1^{(R_N)}]$ to be already decoded at user 1.

a) *Decoding of phase 2:* As a first step, the Q-MAT and ZF data symbols of phase 2 of round N are decoded using successive decoding. Indeed, as we see from (22), the SINR of the Q-MAT data symbol $m[t_{1,2}^{(R_N)}]$ is in the order of $P^{1-\alpha}$, which matches the scaling of the data symbol's rate. Using successive decoding, the ZF data symbol $s_1[t_{1,2}^{(R_N)}]$ is then also decoded.

b) *Decoding the interference in phase 1 (round N):* Receiver 1 then uses the signal received during TS $t_2^{(R_{N+1})}$ (i.e., round $N + 1$) to decode $a_1[t_2^{(R_{N+1})}]$. This is possible because, as seen in (10), the scaling of the SINR of $a_1[t_2^{(R_{N+1})}]$ is $P^{\min(\alpha, 1-\alpha)}$. The content of $a_1[t_2^{(R_{N+1})}]$ depends on whether $\alpha \geq \frac{1}{2}$ or $\alpha \leq \frac{1}{2}$.

- If $\alpha \geq \frac{1}{2}$, user 1 has obtained $\hat{i}_1[t_2^{(R_N)}]$ from $a_1[t_2^{(R_{N+1})}]$.
- If $\alpha \leq \frac{1}{2}$, user 1 has obtained $\hat{n}_1[t_2^{(R_N)}]$ from $a_1[t_2^{(R_{N+1})}]$. To recover the quantized interference $\hat{i}_1[t_2^{(R_N)}]$, it is necessary for user 1 to also decode $Q_{1-\alpha, \alpha} \left(i_1[t_2^{(R_N)}] \right)$ to form $\hat{i}_1[t_2^{(R_N)}]$ in a similar way as in (17).

The term $Q_{1-\alpha, \alpha} \left(i_1[t_2^{(R_N)}] \right)$ can be obtained at user 1 by applying the quantizer $Q_{1-\alpha, \alpha}$ to the received signal $y_1[t_2^{(R_N)}]$. Indeed, following Lemma 1, it holds that

$$\lim_{P \rightarrow \infty} \Pr \left\{ Q_{1-\alpha, \alpha} \left(y_1[t_2^{(R_N)}] \right) = Q_{1-\alpha, \alpha} \left(i_1[t_2^{(R_N)}] \right) \right\} = 1. \quad (23)$$

Thus, in the limit of large P , user 1 can decode $Q_{1-\alpha, \alpha} \left(y_1[t_2^{(R_N)}] \right)$ with probability one, and then combine it with $\hat{n}_1[t_2^{(R_N)}]$ to obtain $\hat{i}_1[t_2^{(R_N)}]$.

In both cases $\alpha \geq \frac{1}{2}$ and $\alpha \leq \frac{1}{2}$, user 1 has obtained $\hat{i}_1[t_2^{(R_N)}]$.

c) *Decoding of the Q-MAT data symbols of phase 1:* By induction, the auxiliary data symbol $a_1[t_2^{(R_N)}]$ is known at user 1 such that its contribution to the received signal can be removed. Furthermore, due to the decoding of phase 2, user 1 has obtained

$$m[t_{1,2}^{(R_N)}] = (\hat{i}_1[t_2^{(R_N)}] \oplus \hat{i}_2[t_1^{(R_N)}]). \quad (24)$$

Using $\hat{i}_1[t_2^{(R_N)}]$, user 1 obtains $\hat{i}_2[t_1^{(R_N)}]$, and thus user 1 has knowledge (up to the noise level) of the following two components

$$\underbrace{\mathbf{h}_1^H[t_1^{(R_N)}] \mathbf{V}[t_1^{(R_N)}] \mathbf{m}[t_1^{(R_N)}] + \mathbf{h}_1^H[t_1^{(R_N)}] \mathbf{v}_2^{\text{ZF}}[t_1^{(R_N)}] a_2[t_1^{(R_N)}]}_{\doteq P} + \underbrace{z_1[t_1^{(R_N)}]}_{\doteq P^\alpha} \quad (25)$$

$$\underbrace{\mathbf{h}_2^H[t_1^{(R_N)}] \mathbf{V}[t_1^{(R_N)}] \mathbf{m}[t_1^{(R_N)}]}_{\doteq P^{1-\alpha}}.$$

By induction, $a_2[t_1^{(R_N)}]$ is assumed to be already decoded at user 1, such that its contribution to the received signal in (25) can be removed. Consequently, user 1 has obtained two signals with a SINR scaling in $P^{1-\alpha}$, and can decode its two destined data symbols contained in $\mathbf{m}[t_1^{(R_N)}]$.

d) Decoding of the auxiliary data symbols of round $N + 1$: The decoding of $a_2[t_1^{(R_{N+1})}]$ follows directly from the definition of the auxiliary data symbol in (18). Indeed, the auxiliary data symbol $a_2[t_1^{(R_{N+1})}]$ is a function of the Q-MAT data symbols $\mathbf{m}[t_1^{(R_N)}]$, which have already been decoded by user 1. It can thus decode $a_2[t_1^{(R_{N+1})}]$, which provides the induction to the next round.

e) Decoding of the ZF data symbols at user 1: User 1 has decoded its Q-MAT data symbols and the auxiliary data symbols needed for the next round. Thus, it remains to show that all the private data symbols destined to user 1 can also be decoded. For $t = t_1^{(R_N)}$ and $t = t_{1,2}^{(R_N)}$, user 1 has decoded the Q-MAT data symbols transmitted. Consequently, user 1 can use successive decoding to decode its destined ZF data symbol $s_1[t_1^{(R_N)}]$ and $s_1[t_{1,2}^{(R_N)}]$ (see (7) and (22)). Yet, the Q-MAT data symbols sent during $t_2^{(R_N)}$ have not been decoded at user 1 (as destined to user 2), and thus successive decoding can not be used directly here. Yet, as user 1 has reconstructed $i_1[t_2^{(R_N)}]$ (up to bounded noise), it can use it to remove the interference generated by the Q-MAT data symbols that were meant for user 2, and can consequently decode its own ZF data symbols (see (10)).

Remark 2. The fact that user 1 is able to remove $i_1[t_2^{(R_N)}]$ up to the noise floor, is the key property of our transmission scheme. This allows to remove all the received signals generated by the Q-MAT data symbols during all TS, thus making possible the ZF transmission with power P^α . \square

The decoding process is illustrated in Fig. 3.

C. Calculation of the DoF

In this particular $K = 2$ setting, one Q-MAT round lasts 3 TS during which 2 Q-MAT data symbols of rate $(1 - \alpha) \log_2(P)$ bits are transmitted to each user. Furthermore, one ZF data symbol of rate $\alpha \log_2(P)$ bits is transmitted to each user during each TS. In the last round, termination is achieved by sending only auxiliary symbols, which induces a small loss in DoF performance. This loss is made negligible by considering a large number of rounds. The resulting

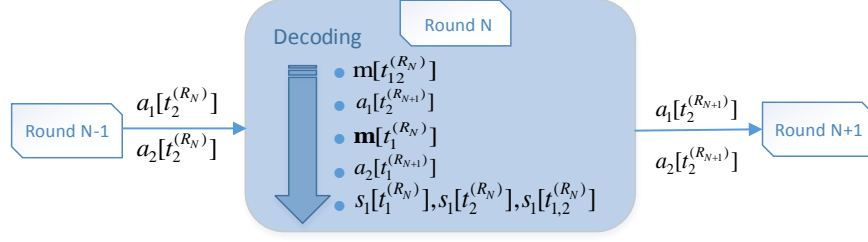


Fig. 3: Illustration of the decoding process for $K = 2$.

DoF is thus

$$\text{DoF}^{\text{Q-MAT}} = \frac{4(1 - \alpha) + 6\alpha}{3}, \quad (26)$$

which concludes the proof for the 2-user case.

IV. K -USER CASE: ENCODING AND TRANSMISSION

We now describe the Q-MAT scheme for an arbitrary number of users. The transmission spans several so-called Q-MAT rounds with each round following the multi-phase structure of MAT [2] with K phases. Phase j aims to communicate order- j (i.e., destined to j users) data symbols. For clarity, our description will trace the description in Section III.C of [2]. In contrast to the original MAT scheme, the order- j data symbols are now digitally encoded such that they are expected to be decoded at all j corresponding RXs.

We proceed to describe the transmission during round N , and assume that the transmissions up to round $N - 1$ have already been concluded.

A. Phase j for $j \in \{1, \dots, K - 1\}$

Phase j consists in the transmission of $(K - j + 1) \binom{K}{j}$ order- j data symbols of rate $(1 - \alpha) \log_2(P)$ bits, and in the process it leads to the generation of $j \binom{K}{j+1}$ order- $(j+1)$ data symbols of rate $(1 - \alpha) \log_2(P)$ bits, as well as auxiliary symbols to be sent during phase j of round $N+1$.

Phase j of round N spans $\binom{K}{j}$ time-slots, each dedicated to a subset \mathcal{S} of users, for all $\mathcal{S} \subset \mathcal{K}$ of size $|\mathcal{S}| = j$ and where we denote by $t_{\mathcal{S}}^{(R_N)}$ the TS dedicated to \mathcal{S} .

1) *Transmission at phase j* : During $t = t_S^{(R_N)}$, the transmit signal is given by

$$\mathbf{x}[t_S^{(R_N)}] = \mathbf{V}[t_S^{(R_N)}] \mathbf{m}[t_S^{(R_N)}] + \sum_{\ell \in \bar{\mathcal{S}}} \mathbf{v}_\ell^{\text{ZF}}[t_S^{(R_N)}] a_\ell[t_S^{(R_N)}] + \sum_{k=1}^K \mathbf{v}_k^{\text{ZF}}[t_S^{(R_N)}] s_k[t_S^{(R_N)}] \quad (27)$$

where

- $\mathbf{m}[t_S^{(R_N)}] \in \mathbb{C}^{K-j+1}$ is a vector containing $(K-j+1)$ Q-MAT order- j data symbols meant for the users in \mathcal{S} . Each symbol carries $(1-\alpha)\log_2(P)$ bits. The first symbol has full power $\mathbb{E} \left[|\{\mathbf{m}[t_S^{(R_N)}]\}_1|^2 \right] \doteq P$, while the others have power $\mathbb{E} \left[|\{\mathbf{m}[t_S^{(R_N)}]\}_i|^2 \right] \doteq P^{1-\alpha}, \forall i \in \{2, \dots, K-j+1\}$. Furthermore, $\mathbf{V}[t_S^{(R_N)}] \in \mathbb{C}^{K \times (K-j+1)}$ is defined as

$$\mathbf{V}[t_S^{(R_N)}] \triangleq \begin{bmatrix} \mathbf{v}_S^{\text{ZF}}[t_S^{(R_N)}] & \mathbf{U}_j \end{bmatrix} \quad (28)$$

where $\mathbf{v}_S^{\text{ZF}}[t_S^{(R_N)}] \in \mathbb{C}^K$ is a unit-norm ZF precoder that is orthogonal to all current-CSI estimates of the channels of the users in $\bar{\mathcal{S}}$, and where $\mathbf{U}_j \in \mathbb{C}^{K \times (K-j)}$ is a randomly chosen, isotropically distributed unitary matrix.

- $a_k[t_S^{(R_N)}] \in \mathbb{C}$, $k \in \bar{\mathcal{S}}$ is an auxiliary data symbol meant for user k , having rate $\min(1-\alpha, \alpha)\log_2(P)$ bits and power $\mathbb{E} \left[|a_k[t_S^{(R_N)}]|^2 \right] \doteq P$.
- $s_k[t_S^{(R_N)}]$, $k \in \mathcal{K}$ is a ZF data symbol meant for user k , having rate $\alpha\log_2(P)$ bits and power $\mathbb{E} \left[|s_k[t_S^{(R_N)}]|^2 \right] \doteq P^\alpha$.

At user $k \in \mathcal{S}$, the received signal for $t = t_S^{(R_N)}$ then takes the form

$$y_k[t_S^{(R_N)}] = \underbrace{\mathbf{h}_k^H[t_S^{(R_N)}] \mathbf{V}[t_S^{(R_N)}] \mathbf{m}[t_S^{(R_N)}]}_{\doteq P} + \underbrace{\mathbf{h}_k^H[t_S^{(R_N)}] \sum_{\ell \in \bar{\mathcal{S}}} \mathbf{v}_\ell^{\text{ZF}}[t_S^{(R_N)}] a_\ell[t_S^{(R_N)}]}_{\doteq P^{1-\alpha}} + \underbrace{z_k[t_S^{(R_N)}]}_{\doteq P^\alpha} \quad (29)$$

where for $k \in \mathcal{K}$, we have

$$z_k[t] \triangleq \underbrace{\mathbf{h}_k^H[t] \mathbf{v}_k^{\text{ZF}}[t] s_k[t]}_{\doteq P^\alpha} + \underbrace{\mathbf{h}_k^H[t] \sum_{\ell=1, \ell \neq k}^K \mathbf{v}_\ell^{\text{ZF}}[t] s_\ell[t]}_{\doteq P^0}. \quad (30)$$

At user $k \in \bar{\mathcal{S}}$, the received signal for $t = t_S^{(R_N)}$ is given by

$$y_k[t_S^{(R_N)}] = \underbrace{\mathbf{h}_k^H[t_S^{(R_N)}] \mathbf{v}_k^{\text{ZF}}[t_S^{(R_N)}] a_k[t_S^{(R_N)}]}_{\doteq P} + \underbrace{i_k[t_S^{(R_N)}]}_{\doteq P^{1-\alpha}} + \underbrace{z_k[t_S^{(R_N)}]}_{\doteq P^\alpha} \quad (31)$$

where we have introduced the short-hand notation $i_k[t_S^{(R_N)}]$ for $k \in \bar{\mathcal{S}}$ as

$$i_k[t_S^{(R_N)}] \triangleq \underbrace{\mathbf{h}_k^H[t_S^{(R_N)}] \mathbf{V}[t_S^{(R_N)}] \mathbf{m}[t_S^{(R_N)}]}_{\doteq P^{1-\alpha}} + \underbrace{\mathbf{h}_k^H[t_S^{(R_N)}] \sum_{\ell \in \bar{\mathcal{S}}, \ell \neq k} \mathbf{v}_\ell^{\text{ZF}}[t_S^{(R_N)}] a_\ell[t_S^{(R_N)}]}_{\doteq P^{1-\alpha}}. \quad (32)$$

For the above, it is easy to see (See for example [30]) that ZF beamforming which uses α -quality current CSIT, reduces the scaling of the interference power by a multiplicative factor of $P^{-\alpha}$.

2) *Generation of new data symbols:* The generation of the data symbols that are to be transmitted in phase $j + 1$ of round N and in phase j of round $N + 1$ from the interference generated during phase j of round N is one of the key ingredients of our scheme. We now consider that the transmissions of phase j have ended for every possible set $\mathcal{S} \subset \mathcal{K}$ for which $|\mathcal{S}| = j$.

a) *Preliminary step: Interference quantization:* At the end of phase j , using delayed CSIT, the TX reconstructs $i_k[t_{\mathcal{S}}^{(R_N)}]$, $\forall k \in \bar{\mathcal{S}}$, which have power that scales as $P^{1-\alpha}$. The next step depends on the value of α .

- If $\alpha \leq \frac{1}{2}$, the TX uses the quantizer of Lemma 1 to obtain

$$\hat{i}_k[t_{\mathcal{S}}^{(R_N)}] = Q_{1-\alpha,\alpha}(i_k[t_{\mathcal{S}}^{(R_N)}]) \quad (33)$$

that comes with a residual quantization noise $n_k[t_{\mathcal{S}}^{(R_N)}]$, which — by design, and directly from the proof of Lemma 1 — has power scaling in P^α . Then the TX quantizes this quantization noise $n_k[t_{\mathcal{S}}^{(R_N)}]$, to get $\hat{n}_k[t_{\mathcal{S}}^{(R_N)}]$, with $\alpha \log_2(P)$ bits, leaving a residual quantization noise that only scales in P^0 (cf. [28]). Finally the TX combines (over the complex numbers) the two quantized estimates, to get a total estimate

$$\hat{i}_k[t_{\mathcal{S}}^{(R_N)}] \triangleq Q_{1-\alpha,\alpha}(i_k[t_{\mathcal{S}}^{(R_N)}]) + \hat{n}_k[t_{\mathcal{S}}^{(R_N)}]. \quad (34)$$

- If $\alpha \geq \frac{1}{2}$, the interference term $i_k[t_{\mathcal{S}}^{(R_N)}]$, $k \in \bar{\mathcal{S}}$ is quantized using any $(1 - \alpha) \log_2(P)$ bit quantizer, to directly give $\hat{i}_k[t_{\mathcal{S}}^{(R_N)}]$ with quantization noise that scales in P^0 (cf. [28]).

b) *Generation of auxiliary data symbols for phase j of round $N + 1$:* The above quantized estimates of the interference will be placed in *auxiliary* data symbols to be transmitted during round $N + 1$. Depending on the value of α , the auxiliary symbol $a_k[t_{\mathcal{S}}^{(R_{N+1})}]$ is loaded as follows

$$\begin{cases} a_k[t_{\mathcal{S}}^{(R_{N+1})}] \leftarrow \hat{n}_k[t_{\mathcal{S}}^{(R_N)}] & , \alpha \leq \frac{1}{2} \\ a_k[t_{\mathcal{S}}^{(R_{N+1})}] \leftarrow \hat{i}_k[t_{\mathcal{S}}^{(R_N)}] & , \alpha \geq \frac{1}{2} \end{cases} \quad (35)$$

where the differentiation into two cases reflects that the rate of the auxiliary data symbol $a_k[t_{\mathcal{S}}^{(R_N)}]$ will be set equal to $\min(1 - \alpha, \alpha) \log_2(P)$ bits. We note that for the first round, all auxiliary data symbols are set to zero.

Remark 3. During each TS, the number of auxiliary data symbols generated matches the number of auxiliary data symbols transmitted, thus ensuring the proper functioning of the scheme. \square

c) Generation of Q-MAT data symbols for phase $j + 1$ of round N : Let us consider an arbitrary set $\mathcal{P} \subset \mathcal{K}$ with $|\mathcal{P}| = j + 1$, and let us denote its elements as

$$\mathcal{P} \triangleq \{p_1, \dots, p_{j+1}\}. \quad (36)$$

The j order- $(j + 1)$ data symbols⁴ $m_\ell[t_{\mathcal{P}}^{(R_N)}], \ell \in \{1, \dots, j\}$ for phase $j + 1$ of round N are then defined as

$$m_\ell[t_{\mathcal{P}}^{(R_N)}] \triangleq \left(\hat{i}_{p_\ell}[t_{\mathcal{P} \setminus p_\ell}^{(R_N)}] \oplus \hat{i}_{p_{\ell+1}}[t_{\mathcal{P} \setminus p_{\ell+1}}^{(R_N)}] \right), \quad \forall \ell \in \{1, \dots, j\}. \quad (37)$$

Each of the above $2j$ components $\{\hat{i}_{p_\ell}[t_{\mathcal{P} \setminus p_\ell}^{(R_N)}], \hat{i}_{p_{\ell+1}}[t_{\mathcal{P} \setminus p_{\ell+1}}^{(R_N)}]\}_{\ell \in \{1, \dots, j\}}$ has already been received at a RX, as some form of interference. In the next phase, the transmitter will recreate the j different linear combinations $m_\ell, \ell \in \{1, \dots, j\}$ and transmit them.

Example 2. For $K = 3$, and $j = 2$, this gives then $\mathcal{P} = \{1, 2, 3\}$ and

$$\begin{aligned} m_1[t_{\{1,2,3\}}^{(R_N)}] &\triangleq (\hat{i}_1[t_{\{2,3\}}^{(R_N)}] \oplus \hat{i}_2[t_{\{1,3\}}^{(R_N)}]), \\ m_2[t_{\{1,2,3\}}^{(R_N)}] &\triangleq (\hat{i}_2[t_{\{1,3\}}^{(R_N)}] \oplus \hat{i}_3[t_{\{1,2\}}^{(R_N)}]). \end{aligned} \quad (38)$$

\square

B. Phase K

Phase K is particular as it does not require to retransmit any data symbols. This is due to the fact that it is a broadcasting phase where a fully common message (meant for all users) is transmitted such that $\mathcal{S} = \mathcal{K}$ and $\bar{\mathcal{S}} = \emptyset$. During the single time-slot $t = t_{\mathcal{K}}^{(R_N)}$ of this phase, the transmitted signal takes the form

$$\mathbf{x}[t] = \mathbf{v}[t]m[t] + \sum_{k=1}^K \mathbf{v}_k^{\text{ZF}}[t]s_k[t] \quad (39)$$

where

- $m[t] \in \mathbb{C}$ is a Q-MAT order- K data symbol, having rate $(1 - \alpha) \log_2(P)$ bits and power $\mathbb{E}[|m[t]|^2] \doteq P$, and $\mathbf{v}[t] \in \mathbb{C}^K$ is a random unit-norm vector.

⁴See further down Remark 1 for a clarification on having to repeat phases to accumulate enough symbols. This is done exactly as in MAT [2], and it is transparent to the scheme and the performance here.

- $s_k[t]$, $k \in \mathcal{K}$ is a ZF data symbol meant for user k , having rate $\alpha \log_2(P)$ bits and power $\mathbb{E}[|s_k[t]|^2] \doteq P^\alpha$.

During the TS $t = t_{\mathcal{K}}^{(R_N)}$, user $k \in \mathcal{K}$ receives

$$y_k[t] = \underbrace{\mathbf{h}_k^H[t] \mathbf{v}[t] m[t]}_{\doteq P} + \underbrace{z_k[t]}_{\doteq P^\alpha}. \quad (40)$$

Remark 4. It is important to note that vector $\mathbf{m}[t_{\mathcal{S}}^{(R_N)}] \in \mathbb{C}^{K-j}$ ($|\mathcal{S}| = j + 1$) contains $K - j$ data symbols. To cover the gap from the fact that we have only generated j order- $(j + 1)$ data symbols $m_\ell[t_{\mathcal{P}}^{(R_N)}]$, $\ell = 1, \dots, j$ ($|\mathcal{P}| = j$), one must simply repeat phase j exactly $\frac{K!}{j}$ times, as in the original MAT scheme [2]. This is automatically accounted for in the DoF calculation, as we will note later. \square

V. SCHEME FOR THE K -USER CASE: DECODING AND DOF

A. Decoding: Proof by Induction

We now consider decoding, and for simplicity assume that all transmissions of all phases and all rounds, have been completed⁵. The proof is done by induction. Before making the inductive statement, we introduce the following set of decoded symbols

$$\mathcal{A}_{j,k}^{(R_N)} \triangleq \left\{ a_\ell[t_{\mathcal{S}}^{(R_N)}] : \forall \mathcal{S} \subset \mathcal{K}, \forall \ell \in \bar{\mathcal{S}}, |\mathcal{S}| = j, k \in \mathcal{S} \right\}. \quad (41)$$

Intuitively, $\mathcal{A}_{j,k}^{(R_N)}$ is the set of all the auxiliary data symbols generated by a transmission during phase j of round N where user k was among the destined users.

For any round N and phase $j < K$, our induction statement is as follows:

If

- (i) Phase $j + 1$ up to phase K of round N have been successfully decoded
- (ii) User k has decoded the set $\mathcal{A}_{j,k}^{(R_N)}$

then

- (i) User k can decode all its destined Q-MAT order- j data symbols and its destined ZF data symbols transmitted during phase j of round N
- (ii) User k can decode the set $\mathcal{A}_{j,k}^{(R_{N+1})}$.

⁵It will become clear that the data symbols of round N can be decoded as soon as round $N + 1$ has ended. Thus, the delay in the decoding of one data symbol does not increase with the number of rounds.

After initializing the auxiliary data symbols of the first round to zero, each user k can decode $\mathcal{A}_{j,k}^{(R_1)}$ as they have a rate equal to zero. Thus we proceed with the inductive step for an arbitrary round N and phase $j < K$, where by induction, it holds that all Q-MAT order- $(j+1)$ data symbols have been successfully decoded at the corresponding users, which means that user k has received the data symbols $m_\ell[t_{\mathcal{T}}^{(R_N)}]$, $\ell = 1, \dots, j$, for any set \mathcal{T} for which $|\mathcal{T}| = j+1$ and for which $k \in \mathcal{T}$.

In addition, still by induction, user k has decoded the set $\mathcal{A}_{j,k}^{(R_N)}$.

d) Decoding of the desired auxiliary data symbols: As a first step, user k uses the signal received during round $N+1$ to decode $a_k[t_{\mathcal{W}}^{(R_{N+1})}]$, $\forall \mathcal{W} \subset \mathcal{K}, |\mathcal{W}| = j, k \in \bar{\mathcal{W}}$. Indeed, it can be seen from (31) that the scaling of the SINR matches the scaling of the rate. Again we differentiate between the cases $\alpha \geq \frac{1}{2}$ and $\alpha \leq \frac{1}{2}$ as the information contained in $a_k[t_{\mathcal{W}}^{(R_{N+1})}]$ which be different.

- If $\alpha \leq \frac{1}{2}$, user k has decoded $\hat{n}_k[t_{\mathcal{W}}^{(R_N)}]$ for every set $\mathcal{W} \subset \mathcal{K}$ for which $k \in \bar{\mathcal{W}}$ and for which $|\mathcal{W}| = j$ (see (35)). To recover the quantized interference $\hat{i}_k[t_{\mathcal{W}}^{(R_N)}]$, it is necessary for user k to obtain $Q_{1-\alpha,\alpha}(i_k[t_{\mathcal{W}}^{(R_N)}])$ (see (34)), and this is achieved by quantizing the received signal $y_k[t_{\mathcal{W}}^{(R_N)}]$ using the quantizer $Q_{1-\alpha,\alpha}$. Indeed, it follows from Lemma 1 that it holds with probability that approaches 1 as the transmit power P increases, that

$$Q_{1-\alpha,\alpha}\left(y_k[t_{\mathcal{W}}^{(R_N)}]\right) = Q_{1-\alpha,\alpha}\left(i_k[t_{\mathcal{W}}^{(R_N)}]\right). \quad (42)$$

- If $\alpha \geq \frac{1}{2}$, user k has decoded directly from the auxiliary data symbols the quantized interference $\hat{i}_k[t_{\mathcal{W}}^{(R_N)}]$ for every set \mathcal{W} with $|\mathcal{A}| = j$ such that $k \in \bar{\mathcal{W}}$ (see (35)).

In both cases, user k has now obtained $\hat{i}_k[t_{\mathcal{W}}^{(R_N)}]$, $\forall \mathcal{W} \subset \mathcal{K}, |\mathcal{W}| = j, k \in \bar{\mathcal{W}}$.

e) Information at user k at the end of phase $j+1$: We now describe what are the data symbols available at user k after successfully decoding phase $j+1$.

Note that for any set $\mathcal{T}, |\mathcal{T}| = j+1$ for which $k \in \mathcal{T}$, we can write $\mathcal{T} = \{k, \mathcal{W}_{\mathcal{T},k}\}$ with $|\mathcal{W}_{\mathcal{T},k}| = j$ and $k \in \bar{\mathcal{W}}_{\mathcal{T},k}$. Thus from the decoding of the auxiliary data symbols during phase j in the previous paragraph, user k knows $\hat{i}_k[t_{\mathcal{W}_{\mathcal{T},k}}^{(R_N)}]$. Using these interference terms in combination with the order- $(j+1)$ data symbols $m_\ell[t_{\mathcal{T}}^{(R_N)}]$, $\ell = 1, \dots, j$, user k is able to obtain all the quantized interference terms forming the order- $(j+1)$ data symbols, i.e., $\hat{i}_\ell[t_{\mathcal{W}_{\mathcal{T},\ell}}^{(R_N)}]$, $\forall \ell \in \mathcal{T}$ with $\ell \neq k$ and $\mathcal{T} = \{\ell, \mathcal{W}_{\mathcal{T},\ell}\}$, where again we note that $k \in \mathcal{W}_{\mathcal{T},\ell}$.

Considering all the decoded order- $(j + 1)$ data symbols, user k has then decoded the set \mathcal{O}_k^j defined as

$$\mathcal{O}_k^j \triangleq \left\{ \left\{ \hat{i}_\ell[t_{\mathcal{T} \setminus \ell}^{(R_N)}] \right\}_{\ell \in \mathcal{T}} : \forall \mathcal{T} \subset \mathcal{K}, |\mathcal{T}| = j + 1, k \in \mathcal{T} \right\}. \quad (43)$$

f) Decoding of the Q-MAT data symbols at user k : Let us now consider an arbitrary user k and an arbitrary set $\mathcal{S} \subset \mathcal{K}, |\mathcal{S}| = j, k \in \mathcal{S}$. We will show that user k is able to decode its destined $K - j + 1$ Q-MAT order- j data symbols in $\mathbf{m}[t_{\mathcal{S}}^{(R_N)}]$. For that purpose, user k needs $K - j + 1$ observations with a SINR scaling in $P^{1-\alpha}$. These observations will be:

$$y_k[t_{\mathcal{S}}^{(R_N)}] \quad (44)$$

$$i_\ell[t_{\mathcal{S}}^{(R_N)}], \forall \ell \in \bar{\mathcal{S}}. \quad (45)$$

Indeed, we can rewrite (up to the noise floor) the above quantized interference terms as $\hat{i}_\ell[t_{\mathcal{T} \setminus \ell}^{(R_N)}]$ with $\mathcal{T} = \{\mathcal{S}, \ell\}$ and $k \in \mathcal{T}$. Thus, it can be seen from (45) that all the quantized interference terms $\hat{i}_\ell[t_{\mathcal{S}}^{(R_N)}], \forall \ell \in \bar{\mathcal{S}}$ are part of \mathcal{O}_k^j . Hence, user k is able to obtain all the observations in (45). Yet, the Q-MAT data symbols are interfered by the auxiliary data symbols $a_\ell[t_{\mathcal{S}}^{(R_N)}]$ (see (29) and (31)), but by induction, these data symbols are known at user k , and thus their interference can be removed. Consequently, user k is able to decode its corresponding Q-MAT data symbols in $\mathbf{m}[t_{\mathcal{S}}^{(R_N)}]$.

g) Decoding of the ZF data symbols at user k : With the Q-MAT order- j data symbols decoded, it is possible for user k to decode the ZF data symbols $s_k[t_{\mathcal{S}}^{(R_N)}]$ for all the time slots $t_{\mathcal{S}}^{(R_N)}$ for which $k \in \mathcal{S}$ ($|\mathcal{S}| = j$). When $k \in \bar{\mathcal{S}}$, user k cannot decode the Q-MAT data symbols transmitted, as it does not have sufficiently many observations. However, the user has decoded $\hat{i}_\ell[t_{\mathcal{S}}^{(R_N)}]$ for all sets $\mathcal{S} \subset \mathcal{K}$ with $|\mathcal{S}| = j$ and $k \in \bar{\mathcal{S}}$. Therefore it can remove the interference created by the Q-MAT data symbols up to the noise floor also for these TS, and then decode its destined ZF data symbol $s_k[t_{\mathcal{S}}^{(R_N)}]$.

h) Decoding of the auxiliary data symbols of round $N + 1$: In order to conclude the inductive step, it remains to prove that it is possible for user k to decode the auxiliary data symbols $a_\ell[t_{\mathcal{S}}^{(R_{N+1})}], \forall \ell \in \bar{\mathcal{S}}, |\mathcal{S}| = j, k \in \mathcal{S}$. This follows directly from the definition of the auxiliary data symbols in (35). Indeed, these auxiliary data symbols at round $N + 1$ are a function of the auxiliary data symbols at round N (i.e., $a_\ell[t_{\mathcal{S}}^{(R_N)}]$ for $\ell \in \bar{\mathcal{S}}$, which are assumed to be known by induction), and the Q-MAT data symbols in $\mathbf{m}[t_{\mathcal{S}}^{(R_N)}]$ which have been decoded

at user k . Therefore, user k is able to decode these auxiliary data symbols. This concludes the inductive step.

B. Calculation of the DoF

We first note that the last round is dedicated to transmitting only auxiliary data symbols, and that — if the number of rounds is sufficiently large — this DoF loss is negligible. Thus focusing on just one round, we note that we have followed closely the structure in different phases of the MAT scheme, so to calculate our DoF we first scale the MAT DoF by a factor of $(1 - \alpha)$ to account for our reduced rate of the Q-MAT data symbols, and then note that during each TS, we additionally send $\alpha \log_2(P)$ bits to each user. Adding these two parts together, provides immediately the sum-DoF expression in Theorem 1, and concludes the proof.

VI. CONCLUSION

The work has provided the first ever communication scheme which, for the general K -user MISO BC, manages to simultaneously and optimally exploit both delayed and imperfect-quality current CSIT. This is achieved by providing a new way of jointly incorporating MAT-type alignment (based on delayed CSIT), and ZF-type separation (using imperfect-quality current CSIT). In addition to resolving a theoretical open problem, the Q-MAT scheme is designed to adapt to CSI considerations that span both timeliness and precision. Interestingly, the Q-MAT scheme can also be seen as a robust precoding scheme which exploits the knowledge of the past to make the transmission less dependent on the current channel state. Indeed, the DoF achieved with the Q-MAT scheme degrades more slowly than schemes from the literature as the CSIT quality coefficient α decreases. Investigating how to leverage the new ideas presented in this work to develop practical robust transmission schemes is an interesting research problem currently under investigation within our group.

APPENDIX

The quantizer Q_{β_1, β_2} quantizes separately the imaginary part and the real part using half the bits for each dimension. Therefore, we consider in the following the quantization of a real

signal y . Let us start by introducing the short-hand notations:

$$\begin{aligned} y_P &\triangleq \sqrt{P^{\beta_1}} y, \\ n_P &\triangleq \sqrt{P^{\beta_2}} n \end{aligned} \quad (46)$$

and denote the points after applying the quantizer as

$$\begin{aligned} \hat{y}_P &\triangleq Q_{\beta_1, \beta_2}(y_P) \\ \hat{y}_P^n &\triangleq Q_{\beta_1, \beta_2}(y_P + n_P). \end{aligned} \quad (47)$$

We will now show how to design the quantizer Q_{β_1, β_2} such that

$$\lim_{P \rightarrow \infty} \Pr\{\hat{y}_P = \hat{y}_P^n\} = 1 \quad (48)$$

and

$$\mathbb{E}[|\hat{y}_P - y_P|^2] \leq P^{\beta_2}. \quad (49)$$

The quantizer Q_{β_1, β_2} uses in fact only $\frac{(\beta_1 - \beta_2)}{2} \log_2(P) - \frac{1}{2} \log_2(\log_2(P))$ bits. This unusual fact that a small fraction of the available bits are not used is a consequence of (48): The sensitivity of the quantization to an additive noise has to be controlled.

Building upon rate distortion theory [28], we compute the Mean Square Error (MSE)-minimizing optimal codebook with the rate $\frac{(\beta_1 - \beta_2)}{2} \log_2(P) - \frac{1}{2} \log_2(\log_2(P))$ bits, which we denote by \mathcal{C}^* . It is well known that the Gaussian source is the hardest to compress [28], such that we can upperbound the distortion achieved, denoted by D^* , by the distortion in the Gaussian case to obtain [28]

$$D^* \leq P^{\beta_1} 2^{-2\left(\frac{(\beta_1 - \beta_2)}{2} \log_2(P) - \frac{1}{2} \log_2(\log_2(P))\right)} \quad (50)$$

$$= \log_2(P) P^{\beta_2} \quad (51)$$

$$\doteq P^{\beta_2}. \quad (52)$$

Using this quantizer allows then to satisfy (49). We will now modify this quantizer in order to also satisfy (48). Specifically, we modify \mathcal{C}^* such that the minimal distance between any two points is at least equal to $\sqrt{\log_2(P) P^{\beta_2}}$. This is done as follows: If two points are closer than $\sqrt{\log_2(P) P^{\beta_2}}$, then we remove one of the two points of the codebook. Trivially, this does not increase the scaling of the MSE, such that (49) remains valid. We denote this modified codebook by $\mathcal{C}_{\beta_1, \beta_2}$. The quantizer Q_{β_1, β_2} then maps the signal y to the nearest point inside $\mathcal{C}_{\beta_1, \beta_2}$ and we denote the set of the real values delimiting the quantization cells by $\mathcal{B}_{\beta_1, \beta_2}$.

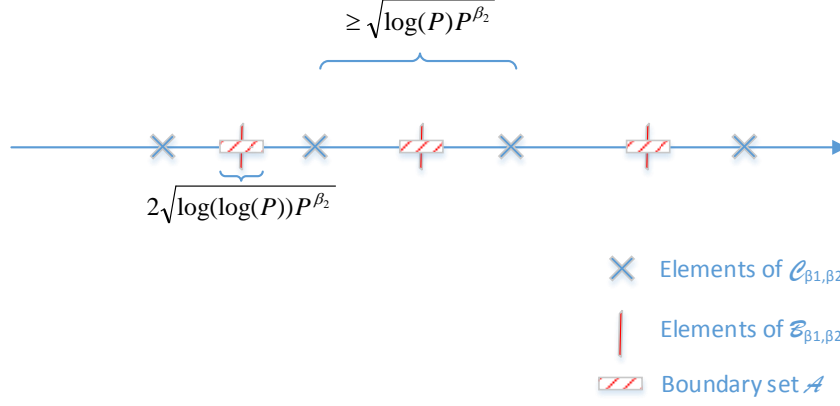


Fig. 4: Illustration of the quantizer design.

We have now finalized the design of the quantizer and it remains solely to show that (48) is indeed satisfied. To prove that result, we start by introducing a set \mathcal{A} containing all the points which are within $\sqrt{\log(\log(P))P^{\beta_2}}$ of the boundary of the quantization cell:

$$\mathcal{A} \triangleq \left\{ x \in \mathbb{R} \mid \min_{q \in \mathcal{B}_{\beta_1, \beta_2}} |x - q| \leq \sqrt{\log(\log(P))P^{\beta_2}} \right\}. \quad (53)$$

To gain insights, the sets $\mathcal{C}_{\beta_1, \beta_2}$, $\mathcal{B}_{\beta_1, \beta_2}$, and \mathcal{A} are illustrated in a toy-example in Fig. 4.

We can then bound the probability that the additive noise n_P leads to a change of quantization point, i.e., $\hat{y}_P \neq \hat{y}_P^n$, as

$$\Pr \{ \hat{y}_P \neq \hat{y}_P^n \} = \Pr \{ \hat{y}_P \neq \hat{y}_P^n | y_P \in \mathcal{A} \} \Pr \{ y_P \in \mathcal{A} \} + \Pr \{ \hat{y}_P \neq \hat{y}_P^n | y_P \notin \mathcal{A} \} \Pr \{ y_P \notin \mathcal{A} \} \quad (54)$$

$$\leq \underbrace{\Pr \{ y_P \in \mathcal{A} \}}_{\triangleq P_1} + \underbrace{\Pr \{ \hat{y}_P \neq \hat{y}_P^n | y_P \notin \mathcal{A} \}}_{\triangleq P_2} \quad (55)$$

where we have denoted by P_1 the first term of (55) and by P_2 the second one. Focusing first on P_2 , we can upperbound it as

$$P_2 \leq \Pr \left\{ |n_P| > \sqrt{\log(\log(P))P^{\beta_2}} \right\} \quad (56)$$

$$\stackrel{(a)}{\leq} \frac{P^{\beta_2}}{\log(\log(P))P^{\beta_2}} \quad (57)$$

$$= \frac{1}{\log(\log(P))} \quad (58)$$

where inequality (a) follows from Chebyshev's inequality.

Turning to P_1 , we can write that

$$P_1 = \Pr \{y_P \in \mathcal{A}\} \quad (59)$$

$$= \Pr \left\{ y \in \frac{\mathcal{A}}{\sqrt{P^{\beta_1}}} \right\} \quad (60)$$

$$= \int_{\frac{\mathcal{A}}{\sqrt{P^{\beta_1}}}} P_Y(y) dy \quad (61)$$

$$\leq \sup_{y \in \mathbb{R}} p_Y(y) \int_{\frac{\mathcal{A}}{\sqrt{P^{\beta_1}}}} dy. \quad (62)$$

We can then upperbound the support of the set \mathcal{A} as its restriction to each quantization cell has the length of $2\sqrt{\frac{\log_2(\log_2(P))P^{\beta_2}}{P^{\beta_1}}}$, for each of the $\sqrt{\frac{P^{\beta_1-\beta_2}}{\log_2(P)}}$ quantization cells. This gives

$$P_1 \leq \sup_{y \in \mathbb{R}} p_Y(y) 2\sqrt{\frac{\log_2(\log_2(P))P^{\beta_2}}{P^{\beta_1}}} \sqrt{\frac{P^{\beta_1-\beta_2}}{\log_2(P)}} \quad (63)$$

$$= \sup_{y \in \mathbb{R}} p_Y(y) 2\sqrt{\frac{\log_2(\log_2(P))}{\log_2(P)}}. \quad (64)$$

Letting P grow to infinity, both P_1 and P_2 tend to zero, which concludes the proof of the lemma.

REFERENCES

- [1] A. G. Davoodi and S. A. Jafar, “Aligned image sets under channel uncertainty: Settling conjectures on the collapse of Degrees of Freedom under finite precision CSIT,” *IEEE Trans. Inf. Theo.*, vol. 62, no. 10, pp. 5603–5618, Oct. 2016.
- [2] M. Maddah-Ali and D. Tse, “Completely stale transmitter channel state information is still very useful,” *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4418–4431, Jul. 2012.
- [3] C. S. Vaze and M. K. Varanasi, “The degrees of freedom region and interference alignment for the MIMO interference channel with delayed CSIT,” *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4396–4417, Jul. 2012.
- [4] M. J. Abdoli, A. Ghasemi, and A. K. Khandani, “On the degrees of freedom of three-user MIMO Broadcast Channel with delayed CSIT,” in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2011.
- [5] N. Lee and R. W. Heath, “Not too delayed CSIT achieves the optimal degrees of freedom,” in *Proc. Allerton Conference on Communication, Control, and Computing (Allerton)*, 2012.
- [6] R. Tandon, S. A. Jafar, S. Shamai (Shitz), and H. V. Poor, “On the synergistic benefits of alternating CSIT for the MISO BC,” *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4106–4128, 2013.
- [7] B. Rassouli, C. Hao, and B. Clerckx, “DoF analysis of the K-user MISO Broadcast Channel with hybrid CSIT,” in *Proc. IEEE International Conference on Communications (ICC)*, 2015.
- [8] X. Yi and D. Gesbert, “Precoding Methods for the MISO Broadcast Channel with Delayed CSIT,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 1–11, 2013.
- [9] A. Vahid, M. A. Maddah-Ali, and A. S. Avestimehr, “Approximate Capacity Region of the MISO Broadcast Channels With Delayed CSIT,” *IEEE Trans. on Commun.*, vol. 64, no. 7, pp. 2913–2924, July 2016.

- [10] M. Kobayashi, S. Yang, D. Gesbert, and X. Yi, "On the degrees of freedom of time correlated MISO Broadcast Channel with delayed CSIT," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2012.
- [11] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, "Degrees of freedom of time correlated MISO Broadcast Channel with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 315–328, Jan. 2013.
- [12] T. Gou and S. Jafar, "Optimal use of current and outdated channel state information: Degrees of freedom of the MISO BC with mixed CSIT," *IEEE Communications Letters*, vol. 16, no. 7, pp. 1084–1087, Jul. 2012.
- [13] J. Chen and P. Elia, "Can imperfect delayed CSIT be as useful as perfect delayed CSIT? DoF analysis and constructions for the BC," in *Proc. Allerton Conference on Communication, Control, and Computing (Allerton)*, 2012.
- [14] J. Chen, S. Yang, and P. Elia, "On the fundamental feedback-vs-performance tradeoff over the MISO-BC with imperfect and delayed CSIT," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2013.
- [15] J. Chen and P. Elia, "Toward the performance versus feedback tradeoff for the two-user MISO Broadcast Channel," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8336–8356, Dec. 2013.
- [16] X. Yi, S. Yang, D. Gesbert, and M. Kobayashi, "The degrees of freedom region of temporally correlated MIMO networks with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 494–514, Jan. 2014.
- [17] J. Chen and P. Elia, "Symmetric two-user MIMO BC with evolving feedback," in *Proc. Information Theory and Applications Workshop (ITA)*, 2014.
- [18] M. Torrellas, A. Agustin, and J. Vidal, "Retrospective Interference Alignment for the MIMO interference Broadcast Channel," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2015.
- [19] B. Rassouli, C. Hao, and B. Clerckx, "DoF analysis of the MIMO broadcast channel with alternating/hybrid CSIT," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1312–1325, Mar. 2016.
- [20] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach," *IEEE Trans. on Commun.*, vol. PP, no. 99, pp. 1–1, 2016.
- [21] C. Hao, B. Rassouli, and B. Clerckx, "Achievable DoF regions of MIMO networks with imperfect CSIT," 2016. [Online]. Available: <http://arxiv.org/abs/1603.07513>
- [22] A. Bracher and M. A. Wigger, "Feedback and partial message side-information on the semideterministic Broadcast Channel," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2015.
- [23] S. Lashgari, R. Tandon, and S. Avestimehr, "Three-user MISO Broadcast Channel: How much can CSIT heterogeneity help?" in *Proc. IEEE International Conference on Communications (ICC)*, 2015.
- [24] O. Shayevitz and M. Wigger, "On the capacity of the discrete memoryless Broadcast Channel with feedback," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1329–1345, Mar. 2013.
- [25] H. Kim, Y. K. Chia, and A. E. Gamal, "A note on the Broadcast Channel with stale state information at the transmitter," *IEEE Trans. Inf. Theory*, vol. 61, no. 7, pp. 3622–3631, July 2015.
- [26] P. de Kerret, X. Yi, and D. Gesbert, "On the degrees of freedom of the K-user time correlated Broadcast Channel with delayed CSIT," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2013.
- [27] Y. Luo, T. Ratnarajah, and A. K. Papazafeiropoulos, "Degrees-of-freedom regions for the K-user MISO time-correlated Broadcast Channel," 2014. [Online]. Available: <http://arxiv.org/abs/1412.1023>
- [28] T. Cover and A. Thomas, *Elements of information theory*. Wiley-Interscience, Jul. 2006.
- [29] P. de Kerret, D. Gesbert, J. Zhang, and P. Elia, "Optimally Bridging the Gap from Delayed to Perfect CSIT in the K-user MISO BC," in *Proc. IEEE Information Theory Workshop (ITW)*, 2016.

- [30] N. Jindal, "MIMO Broadcast Channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.